

Controllability of Linear Positive Systems: An Alternative Formulation[☆]

Yashar Zeinaly^{a,*}, Jan H. van Schuppen^{b,**}, Bart De Schutter^{a,*}

^a*Mekelweg 2, 2628 CD Delft, The Netherlands*

^b*Mekelweg 4, 2628 CD Delft, The Netherlands*

Abstract

An alternative formulation for the controllability problem of single input linear positive systems is presented. Driven by many industrial applications, this formulation focuses on the case where the region of interest is only a subset of positive orthant rather than the entire positive orthant. To this end, we discuss the geometry of controllable subsets and develop numerically verifiable conditions for polyhedrality of controllable subsets. Finally, we provide a method to check for controllability of a target set based on our approach.

Keywords: Linear positive systems, Controllability, Cones, Polyhedral cones

1. Introduction

In this paper, we revisit the “controllability” concept for discrete-time linear positive systems. Motivated by applications underlying positive systems, we will re-define this concept. We will then provide necessary and sufficient conditions for controllability of a certain class of discrete-time linear positive systems.

The concept of positive systems arises in many applications such as econometrics [1], bio-chemical reactors [2, 3], compartmental systems [4, 5], and transportation system [6, 7], to name a few. The variables in such systems represent

[☆]Full draft for review only.

^{*}Delft Center for Systems and Control, Delft University of Technology

^{**}Delft Institute of Applied Mathematics, Delft University of Technology

Email addresses: y.zeinaly@tudelft.nl (Yashar Zeinaly),

J.H.vanSchuppen@tudelft.nl (Jan H. van Schuppen), b.deschutter@tudelft.nl (Bart De Schutter)

growth rates, concentration levels, mass accumulation, or flows, etc. Obviously, variables of this nature can only assume non-negative values. The theory of positive dynamical systems has been developed to deal with this sort of systems. Of particular interest is the theory of linear positive systems [8], which has its roots in the theory of non-negative matrices and in the geometry of cones [9, 10, 11, 12]. While the theory of linear positive systems has overlaps with general theory of linear systems, there are distinct differences between the two. This is due to the fact that linear positive systems are defined over a cone rather than over a linear subspace. Therefore, many properties of linear systems cannot be generalized to linear positive systems without proper treatment. Moreover, some concepts of general linear systems theory might have to be redefined for linear positive systems. One such property is the notion of “controllability” for linear positive systems.

In many industrial applications one might be interested in investigating whether a certain state (e.g., concentration levels) can be reached by applying an appropriate control input. More generally one might be interested in characterizing all states that can be reached from a given initial state using nonnegative control inputs. With respect to this point of view, the alternative approach in this paper is based on the following key problem: *Given a set of states, possibly a singleton, in \mathbb{R}_+^n , can the system initially at rest be steered in finite time to any state of the considered set by applying nonnegative control signal?*

The controllability of discrete-time linear positive systems has been widely studied in the literature. In most of the literature, it has been emphasized that the characterization of controllability for discrete-time linear positive systems takes a very peculiar form, which is very different from its counterpart for discrete-time linear systems [13, 14, 15]. Unlike discrete-time linear systems in \mathbb{R}^n for which reachability is equivalent to reachability in n steps [16], for discrete-time linear positive systems this does not hold and the timing issue becomes very critical, as noted in [14], where they illustrate this using the model of a pharmacokinetic system. However, inspired by the definition of reachability

within the context of linear systems, most papers in the literature investigate and discuss necessary and sufficient conditions under which the positive orthant \mathbb{R}_+^n is reachable. Among others, [15, 17, 18, 19, 20], are some of the significant works that fall in this category. In [17, 15] controllability of discrete-time linear positive systems is characterized using a graph-theoretic approach, and canonical controllability forms are derived as well. The authors of [19] have established a link between positive state controllability and positive input controllability of a related system, which is then used to obtain a controllability criterion. A good survey of similar results is provided in [21, 22]. Controllability results for special classes of 1D and 2D systems are provided in [23].

In this paper we first define and characterize the controllable subsets. Then, in Proposition 2 and Proposition 3, we present necessary and sufficient conditions for polyhedrality of the controllable subsets. Theorem 2 and Theorem 3 provide a numerically verifiable method to check for polyhedrality of the controllable subsets based on spectrum of \mathbf{A} . Finally, in Proposition 6 we propose a method to check for controllability of a given subset of \mathbb{R}_+^n . The rest of this paper is organized as follows. In Section 2, inspired by the aforementioned application domains, we formally introduce our view of the controllability problem. In Section 3, we introduce some notation that will be used in the sequel. A characterization of controllable subsets is then provided in Section 4, and the controllability problem is characterized in Section 5.

2. Problem Formulation

2.1. Classical view

We will now introduce the classical view of the controllability problem as discussed in the literature highlighting that the stated conditions for controllability are often too strict and impractical. Then we will formally introduce our view of the controllability problem arguing why it is more suitable, especially from the application point of view.

Remark 1. Different terminologies have been used for the concept of controllability of linear positive systems in the literature. Investigating whether a state is reachable from the origin has been referred to both as “reachability” and “controllability from the origin.” In this paper, in line with the latter terminology, since we assume the system is initially at rest, we will use controllability to refer to “controllability from the origin.”

In most papers of the literature, the characterization of controllability of linear positive systems is based on the following definition, see [13].

Definition 1. “A positive system is said to be completely reachable if all states $x \geq 0$ are reachable in finite time from the origin, that is, if $X_r = \mathbb{R}_+^n$,” where $X_r = \mathbb{R}_+^n$ denotes the cone of all reachable states in finite time using nonnegative inputs.

The underlying idea behind Definition 1 probably originates from making an analogy to reachability of linear systems. This definition is based on the assumption that the state space is $X = \mathbb{R}_+^n$. But if the system starts at the zero state, then it may not be possible with the existing inputs to reach all states of the system. Therefore the states to be reached may be restricted from the full positive orthant $X = \mathbb{R}_+^n$ to a smaller subset of the positive orthant. Hence the condition of controllability has to be adjusted as described in the remainder of the paper. The following theorem ([13, Th. 27]), states the necessary and sufficient condition for reachability with respect to Definition 1 for the single-input case.

Theorem 1. “A discrete-time positive system is completely reachable if it is possible to reorder its state variables in such a way that the input u directly influences only x_1 , and x_i directly influences x_{i+1} for $i = 1, 2, \dots, n - 1$.”

The results for the multi-input case based on Definition 1 are more involved, but they require that the matrix $[B, AB, \dots, A^k B]$ includes a monomial submatrix of dimension n , for some $k \in \mathbb{N}_+$ [17, 21, 18, 15]. Such conditions are often too strong to be satisfied by most of practical systems. In addition, especially from

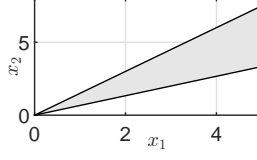


Figure 1: Example 1. The shaded area, associated with K , represents the region of interest for which controllability needs to be checked.

the application point of view, complete reachability according to Definition 1 is not required in most of the cases since many practical positive systems operate in a constrained space, which is a strict subset of \mathbb{R}_+^n and/or we are only interested in reachability of states within a constrained space. For example in economical systems, one would be interested to know whether a certain growth rate can be achieved, which corresponds to checking whether a certain extremal ray is reachable. In bio-chemical reactors, it might be of interest to know whether a set of desired mass concentrations can be achieved by manipulating the inputs (e.g., flow of material). The set of desired concentrations is normally a small subset of \mathbb{R}_+^n .

Example 1. Consider the discrete-time time-invariant linear positive system

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (1)$$

with

$$A = \begin{bmatrix} 4 & 4 \\ 11 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_0 = 0.$$

It is of interest to determine whether the states in the cone $K \subset \mathbb{R}_+^2$, defined by (2) and illustrated by Fig. 1, can be reached in finite time:

$$K : \begin{cases} 3x_1 - 2x_2 \geq 0, \\ 3x_2 - 2x_1 \geq 0, \\ x_1 \geq 0, \quad x_2 \geq 0. \end{cases} \quad (2)$$

Since $K \subset \mathbb{R}_+^2$, in order to answer this question using the classical approach, one

needs to check the reachability of \mathbb{R}_+^2 , which is a very conservative considering the fact that K “occupies” only a small portion of \mathbb{R}_+^2 . It can be verified that

$$[b, Ab, \dots, A^k b] = \begin{bmatrix} 2 & 12 & \cdots \\ 1 & 24 & \cdots \end{bmatrix}$$

does not include a monomial submatrix of dimension 2 for any $k \in \mathbb{N}_+$. Therefore, the conditions of Theorem 1 do not hold and we cannot deduce anything about the reachability of K . Nevertheless, it will be later shown that K is reachable from the origin.

2.2. The approach of this paper

From a practical point of view, the controllability problem boils down to whether it is possible to steer the system at rest to a given target set in finite time; and if this is the case, how long will it take to drive the system there. The controllable subset is defined as the subset of the state set containing those states that are reachable by either a finite or an infinite length nonnegative input signal. That subset of the state set is then a cone. Controllability is then defined as the requirement that the controllable subset contains the target set, which could be different from the positive orthant itself. Therefore the view point has to be changed by focusing on the controllable set, characterizing it, and the determination of conditions which guarantee that a particular subset of the positive orthant is contained in the controllable set. In addition, it will be shown that the controllable set in general does not have a finite characterization.

The notation of a linear positive system is formally defined in Section 3. In this section, the controllable subset is denoted as $\text{Conset}_k(x_0)$, $\text{Conset}_f(x_0)$, or $\text{Conset}_\infty(x_0)$ depending on whether the input sequence contains $k \in \mathbb{N}$ elements, a finite number, or an infinite number of elements.

Consider a discrete time-invariant linear positive system. The controllability problem is then composed of the following subproblems:

1. Characterize the controllable subsets $\text{Conset}_k(\mathbf{x}_0)$, $\text{Conset}_f(\mathbf{x}_0)$, and $\text{Conset}_\infty(x_0)$ for the initial state $x_0 = 0$.

2. Determine whether or not the controllable sets $\text{Conset}_f(x_0)$ and $\text{Conset}_\infty(x_0)$ can be computed in a finite number of steps.
3. Determine conditions on the system such that $\text{Conset}_\infty(x_0) = \mathbb{R}_+^n$.
4. Considering a cone $C_{\text{obj}} \subseteq \mathbb{R}_+^n$ of control objectives or a subset of \mathbb{R}_+^n , determine sufficient and necessary conditions with respect to which the following condition holds: $C_{\text{obj}} \subseteq \text{Conset}_f(0)$.

3. Concepts of Linear Positive Systems

3.1. Positive Real Numbers and Positive Matrices

The reader is informed of the following books on positive real numbers and positive matrices: [9, 24]. Books on positive systems or books with chapters on positive systems include [8, 13, 16, 23]. The reader is assumed to be familiar with the integers, the real numbers, and vector spaces. Denote the set of the integers by \mathbb{Z} , the strictly positive integers by $\mathbb{Z}_+ = \{1, 2, \dots\}$, and the set of the natural numbers by $\mathbb{N} = \{0, 1, 2, \dots\}$. For any $n \in \mathbb{Z}_+$ denote the set of the first n integers and of the first n natural numbers by, respectively, $\mathbb{Z}_n = \{1, 2, \dots, n\}$ and $\mathbb{N}_n = \{0, 1, \dots, n\}$.

The real numbers are denoted by \mathbb{R} , the set of the positive real numbers by $\mathbb{R}_+ = [0, \infty)$, and the set of the strictly positive real numbers by $\mathbb{R}_{s+} = (0, \infty)$. The n -dimensional vector space of tuples of real numbers is denoted by \mathbb{R}^n . The associated field of scalars is the set of the real numbers.

The set of the positive real numbers is a semi-ring. It is closed with respect to addition and with respect to multiplication. But it is not closed with respect to the inverse of addition (subtraction). The set of the strictly positive real numbers is closed with respect to inversion.

Consider the set of n tuples of the positive real numbers \mathbb{R}_+^n , with the set of the positive numbers as the set of scalars. This set is closed with respect to addition but it does not have an inverse with respect to addition. The algebraic structure of $(\mathbb{R}_+, \mathbb{R}_+^n)$ is a semi-ring.

For a finite subset $S \subseteq \mathbb{R}^n$, $K \subseteq \mathbb{R}^n$ is the polyhedral cone generated by S if it consists of all finite nonnegative linear combinations of elements of S . For a matrix $\mathbf{M} \in \mathbb{R}^{n \times m}$, we denote $\text{cone}(\mathbf{M})$ as the cone generated by columns of \mathbf{M} . A *ray* of a cone is a line starting in the vertex of the cone and extending to infinity, and lying on the boundary of the cone. It is called an *extreme ray* if it cannot be written as the convex combination of two other rays. A *polyhedral cone* is a cone for which there exists a finite number of extreme rays such that any vector starting at the vertex of the cone and extending to infinity, is a finite nonnegative linear combination of the extremal rays. A cone which is not polyhedral is also called a *round cone*. Thus, a cone is a round cone if there exists a non-denumerable number of extreme rays. An example of a round cone is the well known ice cream cone which may be found in [9].

For a finite set of complex numbers $S = \{s_1, s_2, \dots, s_k\}$, we denote $\rho(S) = \max_{s \in S} |s|$. For $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\rho(\mathbf{A}) = \rho(\text{spec}(\mathbf{A}))$ is the spectral radius of \mathbf{A} , where $\text{spec}(\mathbf{A})$ denotes the set of its eigenvalues. We define the dominant subset of S as $\sigma^\rho(S) = \{s \in \mathbb{C}, |s| = \rho(S)\}$, and the non-dominant subset as $\sigma^-(S) = \{s \in \mathbb{C}, |s| < \rho(S)\}$. For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, we use $\sigma^\rho(\mathbf{A})$ and $\sigma^-(\mathbf{A})$ as the shorthand notation for $\sigma^\rho(\text{spec}(\mathbf{A}))$ and $\sigma^-(\text{spec}(\mathbf{A}))$, respectively.

A matrix $\mathbf{A} \in \mathbb{R}_+^n$ is reducible if there exists a permutation matrix [25] $\mathbf{S} \in \mathbb{R}_+^{n \times n}$ such that $\hat{\mathbf{A}} = \mathbf{S}^T \mathbf{A} \mathbf{S} = \begin{bmatrix} \mathbf{A}_{11} & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$. An irreducible matrix is the one that is not reducible. A positive real scalar $p \in \mathbb{R}_{s+}$ is always irreducible.

An irreducible matrix $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ is of degree of cyclicity h , with $1 \leq h \leq n$, if $\sigma^\rho(\mathbf{A})$ is of multiplicity of one with $\sigma^\rho(\mathbf{A}) = \{\rho(\mathbf{A}) \exp(i2\pi k/h), k = 0, \dots, h-1\}$ [9, Th. 2.20]. Moreover, if $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ is irreducible with degree of cyclicity h , then $\text{spec}(\mathbf{A})$ is invariant with respect to polar rotations of $2k\pi/h$ for any $k \in \mathbb{Z}$.

3.2. Linear Positive Systems

Definition 2. Define a discrete-time time-invariant linear positive system, with representation

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (3)$$

$$\mathbf{y}(t) = C\mathbf{x}(t), \quad (4)$$

if for any $\mathbf{x}_0 \in \mathbb{R}_+^n$ and for any input function $\mathbf{u} : T \rightarrow \mathbb{R}_+^m$ it holds that the solution of the difference equation (3) is such that $\mathbf{x}(t) \in \mathbb{R}_+^n$ and $\mathbf{y}(t) \in \mathbb{R}_+^p$ both for all $t \in T$. Call $A \in \mathbb{R}^{n \times n}$ the system matrix, $B \in \mathbb{R}^{n \times m}$ the input matrix, and $C \in \mathbb{R}^{p \times n}$ the output matrix.

It is well known that the solution of the difference equation (3) exists and is provided by the formula,

$$\mathbf{x}(t) = A^t \mathbf{x}_0 + \sum_{i=1}^t A^{i-1} B \mathbf{u}(t-i). \quad (5)$$

Denote this relation by the expression $(0, \mathbf{x}_0) \xrightarrow{\mathbf{u}(0:t-1)} (t, \mathbf{x}(t))$.

3.3. Controllable Subsets

Definition 3. Consider a discrete-time time-invariant linear positive system with representation

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{u}(t), \quad \mathbf{x}(0) = \mathbf{x}_0,$$

$$\mathbf{y}(t) = C\mathbf{x}(t).$$

Define the following subsets of the state space: the k -step controllable subset, the finite controllable subset, and the infinite controllable subset, respectively as the sets,

$$\text{Conset}_k(\mathbf{A}, \mathbf{B}; \mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}_+^n \mid \exists \mathbf{u}(0:k-1), (0, \mathbf{x}_0) \xrightarrow{\mathbf{u}(0:k-1)} (k, \mathbf{x})\}, \quad (6)$$

$$k \in \mathbb{Z}^+,$$

$$\text{Conset}_f(\mathbf{A}, \mathbf{B}; \mathbf{x}_0) = \cup_{k=0}^{\infty} \text{Conset}_k(\mathbf{A}, \mathbf{B}; \mathbf{x}_0) \quad (7)$$

$$\text{Conset}_{\infty}(\mathbf{A}, \mathbf{B}; \mathbf{x}_0) = \overline{\text{Conset}_f(\mathbf{A}, \mathbf{B}; \mathbf{x}_0)}, \forall \mathbf{x}_0 \in \mathbb{R}_+^n, \quad (8)$$

where we have used the notation \overline{S} to denote the closure of the set S with respect to the Euclidean topology. If the initial state equals zero, $\mathbf{x}_0 = 0$, then that state is omitted in the notation as in $\text{Conset}_k(\mathbf{A}, \mathbf{B})$.

4. Characterization of the Controllable Subsets

Proposition 1. Consider a discrete-time linear positive system with the system representation (3) with $\mathbf{x}_0 = 0$. The k -step controllable subset, the finite controllable subset, and the infinite controllable subset equal the expressions

$$\text{Conset}_k(\mathbf{A}, \mathbf{B}) = \text{cone}(\text{conmat}_k(\mathbf{A}, \mathbf{B})), \quad (9)$$

$$\text{Conset}_f(\mathbf{A}, \mathbf{B}) = \text{cone}(\mathbf{B} \ \mathbf{A}\mathbf{B} \ \mathbf{A}^2\mathbf{B} \ \dots), \quad (10)$$

$$\text{Conset}_\infty(\mathbf{A}, \mathbf{B}) = \overline{\text{Conset}_f(\mathbf{A}, \mathbf{B})}, \text{ where} \quad (11)$$

$$\text{conmat}_k(\mathbf{A}, \mathbf{B}) = [\mathbf{B} \ \mathbf{A}\mathbf{B} \ \mathbf{A}^2\mathbf{B} \ \dots \ \mathbf{A}^{k-1}\mathbf{B}] \quad (12)$$

PROOF. Using (5) with $\mathbf{x}_0 = 0$ and with any $\mathbf{u} : T \rightarrow \mathbb{R}_+^m$, it follows that

$$\mathbf{x}(k) = [\mathbf{B} \ \mathbf{A}\mathbf{B} \ \dots \ \mathbf{A}^{k-1}\mathbf{B}][\mathbf{u}(k-1)^T \ \mathbf{u}(k-2)^T \ \dots \ \mathbf{u}(0)^T]^T$$

lies in the cone generated by columns of $[\mathbf{B} \ \mathbf{A}\mathbf{B} \ \dots \ \mathbf{A}^{k-1}\mathbf{B}]$ or, equivalently $\text{Conset}_k(\mathbf{A}, \mathbf{B}) = \text{cone}(\text{conmat}_k(\mathbf{A}, \mathbf{B}))$ for any $\mathbf{u} : T \rightarrow \mathbb{R}_+^m$. The characterization of $\text{Conset}_f(\mathbf{A}, \mathbf{B})$ and $\text{Conset}_\infty(\mathbf{A}, \mathbf{B})$ is then derived in a similar manner.

4.1. Polyhedrality of Controllable Subsets

In this section, given an irreducible matrix $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ with degree of cyclicity $1 \leq h \leq n$ and $\mathbf{b} \in \mathbb{R}_+^n$, we first investigate the polyhedrality of $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$, and characterize the necessary and sufficient conditions in terms of $\text{spec}(\mathbf{A})$. We then prove that polyhedrality of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is a special case of polyhedrality of $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ with stricter requirements. In the sequel, it is assumed that $\text{rank}(\text{conmat}_n(\mathbf{A}, \mathbf{b})) = n$. This condition implies that the characteristic polynomial and the minimal polynomial coincide¹. This is a convenient assumption

¹This is due to the fact that \mathbf{A} is similar to the companion matrix of $p_{\mathbf{A}}(\lambda)$ and that for the companion matrix it holds from [25, pp. 146-147] that the characteristic polynomial and the minimal polynomial are equal to $p_{\mathbf{A}}(\lambda)$.

that may be relaxed in a future paper.

Proposition 2. Assume that $A \in \mathbb{R}_+^{n \times n}$ is irreducible with degree of cyclicity $h \in \mathbb{Z}_+$. Define

$$C_{\lim} = \text{cone}(\mathbf{A}_{f,0}\mathbf{b}, \dots, \mathbf{A}_{f,h-1}\mathbf{b})$$

$$\mathbf{A}_{f,i} = \lim_{k \rightarrow \infty} \frac{\mathbf{A}^{kh}}{\rho(\mathbf{A})^{kh}} \mathbf{A}^i, \text{ for } i = 0, \dots, h-1.$$

Then, the infinite controllable subset $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is polyhedral if and only if there exists $k^* \in \mathbb{Z}_+$ such that

$$\text{cone}(\{\text{Conset}_{k^*+1}(\mathbf{A}, \mathbf{b}), C_{\lim}\}) \subseteq \text{cone}(\{\text{Conset}_{k^*}(\mathbf{A}, \mathbf{b}), C_{\lim}\}), \quad (13)$$

or equivalently

$$\mathbf{A}^{k^*} \mathbf{b} \in \text{cone}(\{\text{conmat}_{k^*}(\mathbf{A}, \mathbf{b}), C_{\lim}\}). \quad (14)$$

In (13) and (14), the cone generated by a set of vectors is extended to a cone generated by another cone and a set of vectors.

Remark 2. Note that due to our assumption on \mathbf{A} , $\mathbf{A}^h = \text{diag}(\mathbf{A}_0, \dots, \mathbf{A}_{h-1})$, where $\mathbf{A}_i \in \mathbb{R}_+^{n_i \times n_i}$, $i = 0, \dots, h-1$, is an irreducible matrix of cyclicity $h = 1$ with $\rho(\mathbf{A}_i) = \rho(\mathbf{A})^h$, and where $\sum_{i=0}^{h-1} n_i = n$. Then, due to [9, Th. 2.4.1] the limit matrices $\lim_{p \rightarrow \infty} (\mathbf{A}_i / \rho(\mathbf{A}_i))^p$, $i = 0, \dots, h-1$ exist. Therefore, the matrices $\mathbf{A}_{f,i}$ for $i = 0, \dots, h-1$ exist and, hence, the cone C_{\lim} exists.

PROOF. Sufficiency: We will show that

$$C = \text{cone}(\text{conmat}_{k^*}(\mathbf{A}, \mathbf{b}), \mathbf{A}_{f,0}\mathbf{b} \dots \mathbf{A}_{f,h-1}\mathbf{b})$$

is \mathbf{A} -invariant. Let $\mathbf{x} = \sum_{i=0}^{k^*-1} c_i \mathbf{A}^i \mathbf{b} + \sum_{i=0}^{h-1} c_{f,i} \mathbf{A}_{f,i} \mathbf{b}$ for arbitrary nonnegative coefficients $\mathbf{c} \in \mathbb{R}_+^{k^*}$ and $\mathbf{c}_f \in \mathbb{R}_+^h$. We then have

$$\mathbf{A}\mathbf{x} = \sum_{i=0}^{k^*-1} c_i \mathbf{A}^{i+1} \mathbf{b} + \sum_{i=0}^{h-1} c_{f,i} \mathbf{A} \mathbf{A}_{f,i} \mathbf{b}. \quad (15)$$

Using (14), and noting that

$$\mathbf{A} \mathbf{A}_{f,i} = \mathbf{A}_{f,i+1}, \quad i = 0, \dots, h-2$$

$$\mathbf{A} \mathbf{A}_{f,h-1} = \rho(\mathbf{A})^h \mathbf{A}_{f,0},$$

(15) can be expressed as $\mathbf{Ax} = \sum_{i=0}^{k^*-1} c'_i \mathbf{A}^i \mathbf{b} + \sum_{i=0}^{h-1} c'_{f,i} \mathbf{A}_{f,i} \mathbf{b}$ for some $\mathbf{c}' \in \mathbb{R}_+^{k^*}$ and some $c'_{f,i} \in \mathbb{R}_+^h$. This proves $\mathbf{Ax} \in C$ for any $\mathbf{x} \in C$. Hence, the system trajectory (5) remains in C and $\text{Conset}_\infty(\mathbf{A}, \mathbf{b}) = C$ is polyhedral.

Necessity: Let $\mathbf{x}_\infty = \lim_{k \rightarrow \infty} \frac{\mathbf{A}^k \mathbf{b}}{\rho(\mathbf{A})^k}$. Note that even though \mathbf{x}_∞ does not exist in general, its behavior is characterized by the set of h vectors

$\mathbf{A}_{f,0} \mathbf{b}, \dots, \mathbf{A}_{f,h-1} \mathbf{b}$ [26] (See proof of Lemma 1). Precisely speaking, due to Lemma 1, $\mathbf{x}_\infty \in \text{cone}(\mathbf{A}_{f,0} \mathbf{b}, \dots, \mathbf{A}_{f,h-1} \mathbf{b})$. By the definition of $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ as the closure of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$, and by the above explanation of the vectors \mathbf{x}_∞ , the extremal rays of the polyhedral $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ belong to the sequence $\{\mathbf{A}^k \mathbf{b} \in \mathbb{R}_+, k \in \mathbb{N}\}$ or are extremal rays of the cone, $\text{cone}(\mathbf{A}_{f,0} \mathbf{b}, \dots, \mathbf{A}_{f,h-1} \mathbf{b})$. Again, by the assumption that $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is polyhedral, there exists a finite $k^* \in \mathbb{Z}_+$ such that $\mathbf{A}^{k^*} \mathbf{b} \in \text{cone}(\mathbf{b}, \dots, \mathbf{A}^{k^*-1} \mathbf{b}, \mathbf{A}_{f,0} \mathbf{b}, \dots, \mathbf{A}_{f,h-1} \mathbf{b})$.

It is clear that if (14) is established for an integer $k^* \in \mathbb{Z}_+$, it will hold for any $k \geq k^*$. The smallest integer $k^* \in \mathbb{Z}_+$ satisfying (14) is called the *vertex number*, k_{vert}^∞ , of the controllable subset $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$. Following the steps of the proof of Proposition 2, we can put forward the following corollary.

Corollary 1. *The following statements are equivalent:*

- (a) $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is polyhedral.
- (b) There exists an integer $k_{\text{vert}}^\infty \in \mathbb{Z}_+$ such that $\text{cone}(\mathbf{b} \ \mathbf{Ab} \ \dots \ \mathbf{A}^{k-1} \mathbf{b} \ \mathbf{A}_{f,0} \mathbf{b} \ \dots \ \mathbf{A}_{f,h-1} \mathbf{b})$ is \mathbf{A} -invariant for $k \geq k_{\text{vert}}^\infty$.
- (c) There exists an integer $k_{\text{vert}}^\infty \in \mathbb{Z}_+$ such that for the matrix equation,

$$\begin{aligned} \mathbf{AM} &= \mathbf{MX}, \\ &\exists \text{ a solution } \mathbf{X} \in \mathbb{R}_+^{(k+h) \times (k+h)}, \text{ with } k \geq k_{\text{vert}}^\infty, \text{ where,} \\ \mathbf{M} &= \begin{bmatrix} \mathbf{b} & \mathbf{Ab} & \dots & \mathbf{A}^{k-1} \mathbf{b} & \mathbf{A}_{f,0} \mathbf{b} & \dots & \mathbf{A}_{f,h-1} \mathbf{b} \end{bmatrix}. \end{aligned}$$

Definition 4. A square positive matrix $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ is said to have a *nonnegative*

recursion if it is satisfied that

$$\begin{aligned} \exists n_m \in \mathbb{N}, \exists \{c_0, \dots, c_{n_m-1}\} \in \mathbb{R}_+^{n_m} \text{ such that} \\ \mathbf{A}^{n_m} = \sum_{i=0}^{n_m-1} c_i \mathbf{A}^i, \end{aligned} \quad (17)$$

or equivalently

$$g(\lambda) = \lambda^{n_m} - \sum_{i=0}^{n_m-1} c_i \lambda^i = 0, \quad \forall \lambda \in \text{spec}(\mathbf{A}). \quad (18)$$

In terms of the characteristic polynomial, $p_{\mathbf{A}}(\lambda)$, clearly this implies that

$$g(\lambda) = p_{\mathbf{A}}(\lambda)Q(\lambda), \quad (19)$$

where $Q(\lambda)$ is a polynomial of degree $n_q \geq 0$. It is immediate that

$$n_m = n + n_q \geq n. \quad (20)$$

We are now in the position to state a characterization of Proposition 2 in terms of $\text{spec}(\mathbf{A})$, hence, providing numerically verifiable conditions as to when (14) holds. Let

$$\hat{\mathbf{A}} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{bmatrix},$$

where $\mathbf{S} \in \mathbb{R}^{n \times n}$ is non-singular, and where $\mathbf{A}_1 \in \mathbb{R}^{h \times h}$ with $\text{spec}(\mathbf{A}_1) = \sigma^{\rho}(\mathbf{A})$, $\mathbf{A}_2 \in \mathbb{R}^{(n-h) \times (n-h)}$ with $\text{spec}(\mathbf{A}_2) = \sigma^{-}(\mathbf{A})$. Note that such a decomposition is possible due to the Perron-Frobenius theorem [9, Th. 2.1.4, 2.2.20]. For the pair (\mathbf{A}, \mathbf{b}) of Proposition 2 we then have the following theorem.

Theorem 2. *The following statements are equivalent:*

- (a) *The infinite controllable subset is polyhedral hence there exists an integer $k^* \in \mathbb{Z}_+$ such that*

$$\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b}) = \text{cone}(\text{conmat}_{k^*}(\mathbf{A}, \mathbf{b}) \ \mathbf{A}_{f,0} \mathbf{b} \ \dots \ \mathbf{A}_{f,h-1} \mathbf{b}). \quad (21)$$

Denote the lowest integer for which the above equality holds by $k_{\text{vert}}^{\infty} \in \mathbb{Z}_+$.

- (b) The matrix \mathbf{A}_2 defined above, has a nonnegative recursion.
- (c) If there is a positive $\lambda_r \in \text{spec}(\mathbf{A}_2)$, then
- (c1) $\lambda_r = \rho(\mathbf{A}_2)$.
 - (c2) For any $\lambda \in \sigma^\rho(\mathbf{A}_2)$, $\lambda = \rho(\mathbf{A}_2)\exp(\phi_\lambda 2\pi i)$, where $\phi_\lambda \in \mathbb{Q}$ is a rational number.
 - (c3) $\sigma^\rho(\mathbf{A}_2)$ are simple.
 - (c4) No $\lambda^- \in \sigma^-(\mathbf{A}_2)$ has a polar angle which is an integer multiple² of $2\pi/Mh$.

The proof is established based on a fundamental result [27, Th. 5] on nonnegative recursion, which is also quoted in Appendix A.

PROOF. (a) \Rightarrow (b) \Rightarrow (c): Since $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is polyhedral, according to Corollary 1, there is a sufficiently large $k \geq n - h$ such that the equation

$$\mathbf{A}(\mathbf{b} \mathbf{A} \mathbf{b} \dots \mathbf{A}^{k-1} \mathbf{b} \mathbf{A}_{f,0} \dots \mathbf{A}_{f,h-1}) = (\mathbf{b} \mathbf{A} \mathbf{b} \dots \mathbf{A}^{k-1} \mathbf{b} \mathbf{A}_{f,0} \mathbf{b} \dots \mathbf{A}_{f,h-1} \mathbf{b}) \mathbf{X} \quad (22)$$

has a solution $\mathbf{X} \geq 0$. It can be easily verified using (14)-(16) that

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_3 & \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & \dots & 0 & \alpha_1 \\ 0 & 1 & & 0 & \alpha_2 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & \alpha_k \end{bmatrix}, \quad (23)$$

$$\mathbf{X}_2 = \begin{bmatrix} 0 & 0 & \dots & 0 & \rho(A)^h \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} 0 & 0 & \dots & 0 & \beta_0 \\ 0 & 0 & \dots & 0 & \beta_1 \\ 0 & 0 & & 0 & \beta_2 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & 0 & 0 & \beta_{h-1} \end{bmatrix}. \quad (24)$$

²Note that $\sigma^\rho(\mathbf{A}_2) \subseteq \{\lambda \in \mathbb{C} | \lambda = \rho(\mathbf{A}_2)\exp(2k\pi i/(Mh)), k = 0, \dots, Mh - 1\}$. See Lemma 2 for details.

constitutes a solution, where $\mathbf{X}_1 \in \mathbb{R}_+^{k \times k}$, $\mathbf{X}_2 \in \mathbb{R}_+^{h \times h}$, and $\mathbf{X}_3 \in \mathbb{R}_+^{h \times k}$. Let $p_{\mathbf{X}_1}(\lambda) = \det(\lambda \mathbf{I} - \mathbf{X}_1)$ and $p_{\mathbf{X}_2}(\lambda) = \det(\lambda \mathbf{I} - \mathbf{X}_2)$. Since, by assumption, $k \geq n - h$ and $\text{rank}(\text{conmat}_n(\mathbf{A}, \mathbf{b})) = n$, due to [28, Lemma 3.10], $p_{\mathbf{A}}(\lambda)$ divides $p_{\mathbf{X}}(\lambda) = p_{\mathbf{X}_1}(\lambda)p_{\mathbf{X}_2}(\lambda) = (\lambda^h - \rho(\mathbf{A})^h)(\lambda^k - \alpha_{k-1}\lambda^{k-1} - \dots - \alpha_0)$. Since \mathbf{A} is irreducible with degree of cyclicity h , $p_{\mathbf{A}}(\lambda)$ can be expressed as $p_{\mathbf{A}}(\lambda) = p_{\mathbf{A}_1}(\lambda)p_{\mathbf{A}_2}(\lambda) = (\lambda^h - \rho(\mathbf{A})^h)p_{\mathbf{A}_2}(\lambda)$. Therefore, $p_{\mathbf{A}_2}(\lambda)$ divides $p_{\mathbf{X}_2}(\lambda)$, which, due to statements (A) and (B) of Theorem 5 in Appendix A, proves \mathbf{A}_2 has a nonnegative recursion of the form $\mathbf{A}_2^{k^*} - \alpha_{k^*-1}\mathbf{A}_2^{k^*-1} - \dots - \alpha_0\mathbf{I} = 0$ for some $n - h \leq k^* \leq k$ and for some $\alpha \in \mathbb{R}_+^{k^*}$. Assume \mathbf{A}_2 has a positive eigenvalue. Since \mathbf{A}_2 satisfies a nonnegative recursion, the statements then (C1-C4) in (C) of Theorem 5 hold for $p_{\mathbf{A}_2}(\lambda)$. It is straightforward to check that this implies that (c1)-(c4) holds³.

(c) \Rightarrow (b) \Rightarrow (a): Assume \mathbf{A}_2 has a positive eigenvalue. We need to prove that statements (c1)-(c4) imply a nonnegative recursion for \mathbf{A}_2 of the form $\mathbf{A}_2^{k^*} - \alpha_{k^*-1}\mathbf{A}_2^{k^*-1} - \dots - \alpha_0\mathbf{I} = 0$, for $k^* \geq n - h$ and $\alpha \in \mathbb{R}_+^{k^*}$, and that, in turn, implies polyhedrality of the infinite controllable subset.

First we show that the statements (c1)-(c4) imply the statements (C1)-(C4) of Theorem 5. The statement $\lambda_r \in \sigma^\rho(\mathbf{A}_2)$ implies (C1) of Theorem 5. The requirement of all $\lambda \in \sigma^\rho(\mathbf{A}_2)$ having a rational polar phase implies (C2). The requirement of all $\lambda \in \sigma^\rho(\mathbf{A}_2)$ being simple implies (C3), and (C4) is implied from $\sigma^-(\mathbf{A}_2)$ including no eigenvalue with polar phase $2\pi m/Mh$ for any $m \in \mathbb{Z}$. Next, invoking the equivalence between (C) and (B) of Theorem 5 for $p_{\mathbf{A}_2}(\lambda)$, one can observe that there is a polynomial $Q(\lambda)$ of positive degree such that

$$g(\lambda) = p_{\mathbf{A}_2}(\lambda)Q(\lambda) = \lambda^{k^*} - \alpha_{k^*-1}\lambda^{k^*-1} - \dots - \alpha_0 = 0, \quad (25)$$

for $k^* \geq n - h$ and $\alpha \in \mathbb{R}_+^{k^*}$. It follows from (17) that \mathbf{A}_2 has a nonnegative recursion, which results in (b).

Given (b), there exists a polynomial $g(\lambda)$ of degree $k^* \geq n - h$ satisfying

³Condition $\lambda_r \in \sigma^\rho(\mathbf{A}_2)$ follows from (C1) of Theorem 5, and conditions (c2) and (c3) are, respectively, direct result of (C2) and (C3). Finally, (c4) is implied from (C4) using Lemma 2.

(25), from which one concludes that $p_{\mathbf{A}}(\lambda) = p_{\mathbf{A}_1}(\lambda)p_{\mathbf{A}_2}(\lambda)$ divides $h(\lambda) = p_{\mathbf{A}_1}(\lambda)g(\lambda) = (\lambda^h - \rho(\mathbf{A})^h)(\lambda^{k^*} - \alpha_{k^*-1}\lambda^{k^*-1} - \dots - \alpha_0)$. Now consider the equation $\mathbf{A}\mathbf{M} = \mathbf{M}\mathbf{X}$ with $\mathbf{M} = [\mathbf{b} \ \mathbf{A}\mathbf{b} \ \dots \ \mathbf{A}^{k^*-1}\mathbf{b} \ \mathbf{A}_{f,0}\mathbf{b} \ \dots \ \mathbf{A}_{f,h-1}\mathbf{b}]$, where $\mathbf{X} \in \mathbb{R}^{(n+k^*) \times (n+k^*)}$ is an unknown matrix. Since $\text{conmat}_{k^*}(\mathbf{A}, \mathbf{b})$ is full rank by assumption and $k^* \geq n - h$, \mathbf{M} is as well of full rank. Then, it is known from [28, Lemma 10] that $p_{\mathbf{A}}(\lambda)$ divides $p_{\mathbf{X}}(\lambda)$. Hence, we can choose \mathbf{X} such that $p_{\mathbf{X}}(\lambda) = h(\lambda)$. A possible choice of \mathbf{X} , having substituted k^* for k , is then given by (23)-(24). It is clear from (23)-(24) that \mathbf{X} admits a nonnegative solution. Based on Corollary 1, this implies that $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$ is polyhedral.

Remark 3. For a polyhedral $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$ the following can be observed:

- (a) Due to (20) and from the second part of the proof of Theorem 2 the vertex number of $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$, k_{vert}^{∞} , is at least $n - h$, which implies $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$ has at least n generators. It has exactly n generators (i.e., is simplicial) if and only if $p_{\mathbf{A}_2}(\lambda)$ has non-positive coefficients.
- (b) In the view of Lemma 1, $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$ can be expressed as $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b}) = \text{cone}(\text{conmat}_{k_{\text{vert}}}(\mathbf{b}, \mathbf{A}) \ \mathbf{v}_{f,0} \ \dots \ \mathbf{v}_{f,h-1})$, where $\mathbf{v}_{f,0}, \dots, \mathbf{v}_{f,h-1}$ are the h distinct nonnegative eigenvectors of \mathbf{A}^h associated with the eigenvalue $\rho(\mathbf{A})^h$.

Example 2 (polyhedral $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$). Consider the discrete-time linear time-invariant nonnegative system (3) with system matrices

$$\mathbf{A} = \begin{bmatrix} 0.9727 & 0 & 0.0263 \\ 0.0388 & 0.1273 & 0.2156 \\ 0 & 3.4497 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad (26)$$

where \mathbf{A} is primitive, i.e., is irreducible with degree of cyclicity $h = 1$. We have $\text{spec}(\mathbf{A}) = \{1, 0.9, -0.8\}$. We can assume $\mathbf{A}_1 = 1$, and $\mathbf{A}_2 = \text{diag}(0.9, -0.8)$. Using Theorem 2, it is immediate that conditions (c1) and (c2) hold as $\lambda = 0.9$ is a simple eigenvalue of \mathbf{A}_2 , which equals the spectral radius of \mathbf{A}_2 . Condition (c1) hold as well since the polar angle of $\lambda = -0.8$ is not a integer multiple of the polar angle of $\lambda = 0.9$. Hence, it can be concluded that the infinite controllable subset $\text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$ is polyhedral. We can also conclude that

\mathbf{A}_2 has a nonnegative recursion, which is readily verified as $p_{\mathbf{A}_2}(\lambda) = \lambda^2 - 0.1\lambda - 0.72$. Fig. 2 illustrates the growth of $\text{Conset}_k(\mathbf{A}, \mathbf{b})$. It can be observed that $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is not polyhedral since the cone keeps growing for increasing values of k . Its closure is, however, polyhedral as shown in Fig. 2d.

Example 3 (non-polyhedral $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$). Consider the discrete-time linear time-invariant nonnegative system (3) with system matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0.5 \\ 0 & 0.4 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad (27)$$

where \mathbf{A} has degree of cyclicity $h = 1$. The spectrum of \mathbf{A} is $\text{spec}(\mathbf{A}) = \{-1.05, 0.7116, 1.3383\}$. One can assume $\mathbf{A}_1 = 1.3383$ and $\mathbf{A}_2 = \text{diag}(-1.05, 0.7116)$. It is immediate that condition (c1) of Theorem 2 is not satisfied as $0.7116 \neq \rho(\mathbf{A}_2)$. Therefore, based on this theorem, $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is not polyhedral. This is illustrated by Fig. 3d, from which it is clear that $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ is approaching a round cone as introduced in Section 3.1.

Now we will investigate polyhedrality of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$. Consider the following proposition. We will show that this implies stricter conditions on $\text{spec}(\mathbf{A})$ and that a more conservative version of Theorem 2 applies.

Proposition 3. The finite controllable subset $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is polyhedral if and only if there exists a positive integer $k^* \in \mathbb{Z}_+$ such that

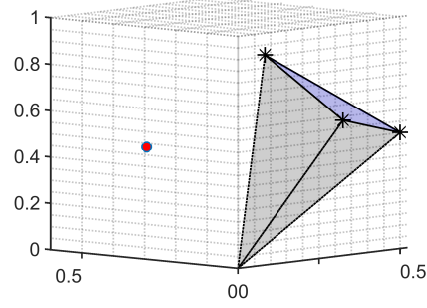
$$\text{Conset}_{k^*+1}(\mathbf{A}, \mathbf{b}) \subseteq \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b}), \quad (28)$$

or equivalently,

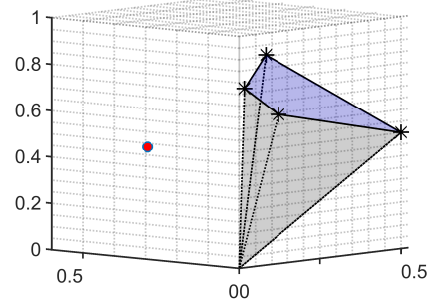
$$\mathbf{A}^{k^*} \mathbf{b} \in \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b}). \quad (29)$$

PROOF. Sufficiency: If $\mathbf{A}^{k^*} \mathbf{b} \in \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b})$ it follows immediately from (5) that $\mathbf{x}(t) \in \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b})$ for any $t \geq k^*$. Hence, based on (10) in Proposition 1, $\text{Conset}_f(\mathbf{A}, \mathbf{b}) = \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b})$.

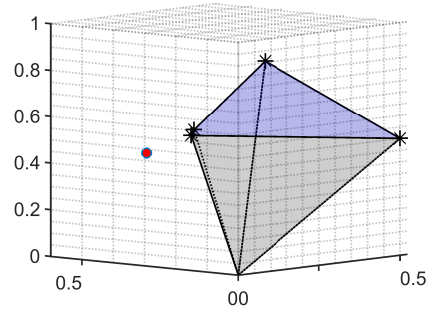
Necessity: if $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is a polyhedral cone, since its generators are of the



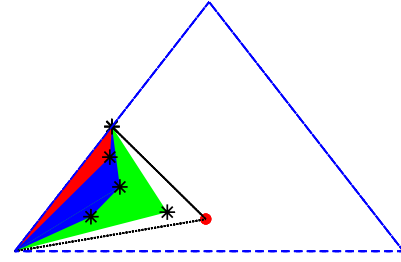
(a) $\text{Conset}_3(\mathbf{A}, \mathbf{b})$



(b) $\text{Conset}_8(\mathbf{A}, \mathbf{b})$

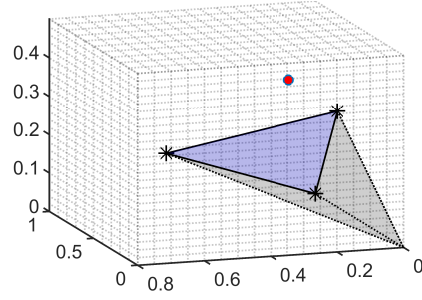


(c) $\text{Conset}_{19}(\mathbf{A}, \mathbf{b})$

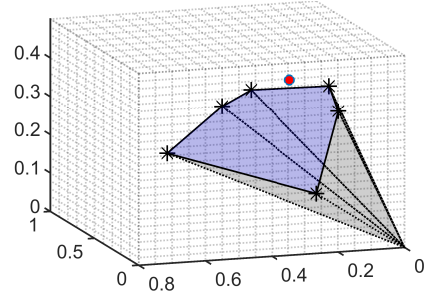


(d) $\text{Conset}_k(\mathbf{A}, \mathbf{b})$, $k=3$ (red), 8 (blue and red), 19 (green, blue and red) and $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ (the triangle with the red vertex)

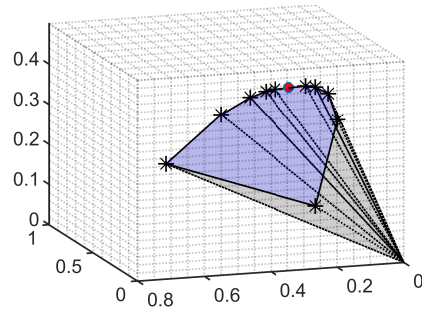
Figure 2: a,b,c: the growth of controllable cone $\text{Conset}_k(\mathbf{A}, \mathbf{b})$ of example 2 for different values of k , where generators of the cone are marked by asterisks, and the Frobenius eigenvector is marked by a red dot. d: The growth of controllable cone mapped on the 3-dimensional simplex $S = \{\mathbf{x} \in \mathbb{R}_+^3 \mid \mathbf{1}^T \mathbf{x} = 1\}$.



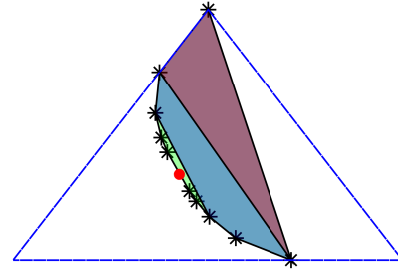
(a) $\text{Conset}_k(\mathbf{A}, \mathbf{b})$, $k = 3$



(b) $\text{Conset}_k(\mathbf{A}, \mathbf{b})$, $k = 6$



(c) $\text{Conset}_k(\mathbf{A}, \mathbf{b})$, $k = 10$



(d) $\text{Conset}_k(\mathbf{A}, \mathbf{b})$, $k = 3$ (red region), $k = 6$ (red and blue regions), $k = 10$ (red, blue and green regions). $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$ approaches a “round cone”.

Figure 3: a,b,c: the growth of controllable cone $\text{Conset}_k(\mathbf{A}, \mathbf{b})$ of example 3 for different values of k , where generators of the cone are marked by asterisks, and the Frobenius eigenvector is marked by a red dot. d: The growth of controllable cone mapped on the 3-dimensional simplex $S = \{\mathbf{x} \in \mathbb{R}_+^3 \mid \mathbf{1}^T \mathbf{x} = 1\}$.

form $\mathbf{A}^k \mathbf{b}$, $k \in \mathbb{Z}_+$, and since a polyhedral cone has a finite number of generators, there must exist a finite $k^* \in \mathbb{Z}_+$ for which $\mathbf{A}^{k^*} \mathbf{b} \in \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b})$.

The smallest k^* for which (29) holds is referred to as the *vertex number*, k_{vert} , of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$. Note that (29) also implies that

$$\text{cone}(\mathbf{A}_{f,0} \mathbf{b} \dots \mathbf{A}_{f,h-1} \mathbf{b}) \subset \text{Conset}_{k_{\text{vert}}}(\mathbf{A}, \mathbf{b}), \quad (30)$$

which is clearly a restriction on (14). Based on the proof of Proposition 3, one can derive the following corollary.

Corollary 2. *The following statements regarding polyhedrality of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ are equivalent:*

- (a) $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is polyhedral.
- (b) There exists an integer $k_{\text{vert}} \in \mathbb{Z}_+$ such that $\text{cone}(\mathbf{b} \mathbf{A} \mathbf{b} \dots \mathbf{A}^k \mathbf{b})$ is \mathbf{A} -invariant for any $k \geq k_{\text{vert}}$.
- (c) There exists an integer $k_{\text{vert}} \in \mathbb{Z}_+$ such that for the matrix equation,

$$\begin{aligned} \mathbf{A}[\mathbf{b} \mathbf{A} \mathbf{b} \dots \mathbf{A}^{k-1} \mathbf{b}] &= [\mathbf{b} \mathbf{A} \mathbf{b} \dots \mathbf{A}^{k-1} \mathbf{b}] \mathbf{X}, \\ &\exists \text{ a solution } \mathbf{X} \in \mathbb{R}_+^{(k) \times (k)}, \text{ with } k \geq k_{\text{vert}}. \end{aligned}$$

- (d) Based on (30) and Lemma 1, there exists an integer $k_{\text{vert}} \in \mathbb{Z}_+$ such that $\text{cone}(\mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1}) \subset \text{Conset}_k(\mathbf{A}, \mathbf{b})$ for any $k \geq k_{\text{vert}}$.

Now, a decomposition of \mathbf{A} is introduced that will be used for stating the next theorem. Given $\mathbf{A} \in \mathbb{R}_+^{n \times n}$, consider $\mathbf{A}_1 \in \mathbb{R}$ and $\mathbf{A}_2 \in \mathbb{R}^{(n-1) \times (n-1)}$, where $\text{spec}(\mathbf{A}_1) = \rho(\mathbf{A})$ and $\text{spec}(\mathbf{A}_2) = \text{spec}(\mathbf{A}) \setminus \{\rho(\mathbf{A})\}$. The decomposition of \mathbf{A} into \mathbf{A}_1 and \mathbf{A}_2 is then given by $\mathbf{A} = \mathbf{S} \text{diag}(\mathbf{A}_1, \mathbf{A}_2) \mathbf{S}^{-1}$, where $\mathbf{S} \in \mathbb{R}^{n \times n}$ is non-singular. Note that such a decomposition is possible due to the Perron-Frobenius theorem [9, Th. 2.1.4, 2.2.20]. With such decomposition of \mathbf{A} at hand, the following theorem provides necessary and sufficient conditions on $\text{spec}(\mathbf{A})$ for polyhedrality of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$. These conditions turn out to be a conservative version of those of Theorem 2.

Theorem 3. *The following statements are equivalent:*

- (a) *The finite controllable subset is polyhedral and hence there exists an integer $k^* \in \mathbb{Z}_+$, $k^* \geq k_{\text{vert}}$ such that $\text{Conset}_f(\mathbf{A}, \mathbf{b}) = \text{Conset}_{k^*}(\mathbf{A}, \mathbf{b})$.*
- (b) *\mathbf{A} has a nonnegative recursion.*
- (c) *\mathbf{A}_2 does not have any positive eigenvalue.*

PROOF. (a) \Rightarrow (b) \Rightarrow (c): Based on Corollary 2 with $k \geq n$ we obtain

$$\mathbf{A}(\text{conmat}_k(\mathbf{A}, \mathbf{b})) = (\text{conmat}_k(\mathbf{A}, \mathbf{b}))\mathbf{X},$$

where $\mathbf{X} \in \mathbb{R}_+^{k \times k}$ is given by

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \alpha_0 \\ 1 & 0 & \cdots & 0 & \alpha_1 \\ 0 & 1 & & 0 & \alpha_2 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & \alpha_{k-1} \end{bmatrix}.$$

Since, by assumption, $\text{conmat}_n(\mathbf{A}, \mathbf{b})$ is full rank and $k \geq n$, there exists [28, Lemma 3.10] a polynomial $Q(\lambda)$ of nonnegative degree such that $p_{\mathbf{A}}(\lambda)Q(\lambda) = p_{\mathbf{X}}(\lambda) = \lambda^k - \alpha_{k-1}\lambda^{k-1} - \cdots - \alpha_1\lambda - \alpha_0$, which, in the view of Definition 4, proves that \mathbf{A} has a nonnegative recursion. Noting that (b) is equivalent to condition (B) of Theorem 5 ([27, Th. 5]), all conditions (C1)-(C4) are then fulfilled. In particular, (C4) holds as conditions (C1)-(C3) are already satisfied for a nonnegative irreducible matrix due to the Perron-Frobenius theorem [9, Th. 2.1.4, 2.2.20]. Condition (C4) requires that no eigenvalue $\lambda^- \in \sigma^-(\mathbf{A})$ has a polar angle of $2\pi k/h$ for $k = 0, \dots, h-1$. Since $\text{spec}(\mathbf{A})$ is invariant under a polar rotation of $2\pi m/h$ for any $m \in \mathbb{Z}$, no $\lambda^- \in \sigma^-(\mathbf{A})$ is then positive. Noting that for an irreducible matrix, $(\sigma^\rho(\mathbf{A}) \setminus \{\rho(\mathbf{A})\}) \cap \mathbb{R}_+ = \emptyset$ and that $\text{spec}(\mathbf{A}_2) = \sigma^-(\mathbf{A}) \cup \sigma^\rho(\mathbf{A}) \setminus \{\rho(\mathbf{A})\}$, one concludes that \mathbf{A}_2 has no positive eigenvalue.

(c) \Rightarrow (b) \Rightarrow (a): Given (c), we have $\text{spec}(\mathbf{A}_2) \cap \mathbb{R}_+ = \emptyset$. For an irreducible matrix it holds that $(\sigma^\rho(\mathbf{A}) \setminus \{\rho(\mathbf{A})\}) \cap \mathbb{R}_+ = \emptyset$. Since $\text{spec}(\mathbf{A}_2) = \sigma^-(\mathbf{A}) \cup (\sigma^\rho(\mathbf{A}) \setminus \{\rho(\mathbf{A})\})$, it follows that $\sigma^-(\mathbf{A}) \cap \mathbb{R}_+ = \emptyset$, from which it can

be immediately concluded that $\exists \lambda \in \sigma^-(\mathbf{A})$, $\lambda = |\lambda| \exp(i2\pi m/h)$ for any $m \in \mathbb{Z}$. Hence, we established that (C4) of Theorem 5 ([27, Th. 5]) holds for $p_{\mathbf{A}}(\lambda)$. Moreover, statements (C1)-(C3) as well hold for $p_{\mathbf{A}}(\lambda)$ as \mathbf{A} is irreducible. Therefore, due to (B) of Theorem 5, there exists a polynomial $Q(\lambda)$ of nonnegative degree, such that $p_{\mathbf{A}}(\lambda)Q(\lambda) = \lambda^{k^*} - \alpha_{k^*-1}\lambda^{k^*-1} - \dots - \alpha_1\lambda - \alpha_0$, where $k^* \geq n$ and $\alpha_i \geq 0$, $i = 0, 1, \dots, k^* - 1$. This proves that \mathbf{A} has a nonnegative recursion based on Definition 4. Then, (a) immediately follows as $\mathbf{A}^{k^*} \mathbf{b} = \sum_{i=0}^{k^*-1} \alpha_i \mathbf{A}^i \mathbf{b}$.

Remark 4. Note that since $\deg(Q(\lambda)) \geq 0$, k_{vert} of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is at least n , and it equals n if and only if $p_{\mathbf{A}}(\lambda) = \lambda^n - \alpha_{n-1}\lambda^{n-1} - \dots - \alpha_1\lambda - \alpha_0$ with $\alpha_i \geq 0$, $i = 0, \dots, n-1$. Hence $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is a simplicial cone (i.e., has n generators) if and only if the characteristic polynomial of \mathbf{A} has non-positive coefficients. One such matrix is a cyclic matrix with cyclicity index $h = n$ as $p_{\mathbf{A}}(\lambda) = \lambda^n - \rho(\mathbf{A})\lambda^{n-1}$.

Comparing Theorem 2 to Theorem 3 reveals that the latter is a restricted version of the former. For example, Theorem 2b requires a part of $\mathbf{A}(\mathbf{A}_2)$ to have a nonnegative recursion while Theorem 3b requires \mathbf{A} to have a nonnegative recursion.

Example 4 (polyhedral $\text{Conset}_f(\mathbf{A}, \mathbf{b})$). Consider the discrete-time linear time-invariant nonnegative system (3) with system matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 1.6333 & 1.1049 & 0 \\ 23.5667 & 6.0944 & 0 & 0 \\ 0 & 0 & 1.1225 & 1.0672 \\ 0 & 1.6611 & 0 & 0.7830 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad (31)$$

where \mathbf{A} is irreducible with degree of cyclicity $h = 1$. It can be verified that $\text{spec}(\mathbf{A}) = \{10, -4, 1+i, 1-i\}$. One can recognize that no eigenvalue of $\mathbf{A}_2 = \text{diag}(-4, 1+i, 1-i)$ is positive. Therefore, condition (c3) of Theorem 3 holds and it follows that \mathbf{A} has a nonnegative recursion. In fact, it can be verified that in this case it holds that $\mathbf{A}^6 = 166.7569\mathbf{I}_4 + 16.1434\mathbf{A} + 39.7036\mathbf{A}^4 + 6.0262\mathbf{A}^5$,

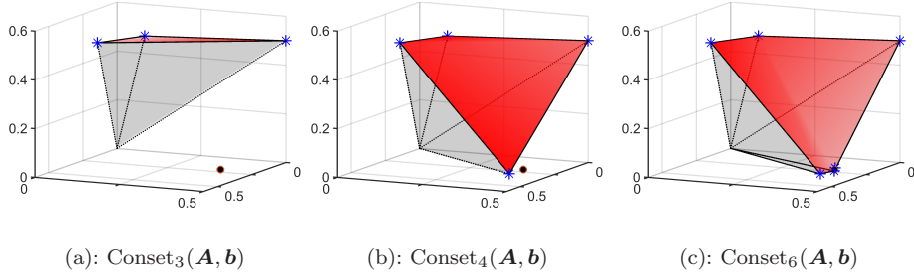


Figure 4: Example 4: growth of the controllable cone mapped on the 3-dimensional simplex $S = \{\mathbf{x} \in \mathbb{R}_+^3 \mid \mathbf{1}^T \mathbf{x} = 1\}$; the generators of the cone and the Frobenius eigenvector are, respectively, marked by asterisks and a dot.

where \mathbf{I}_4 denotes the identity matrix of dimension 4×4 . In addition, we can conclude that $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ is polyhedral with $k_{\text{vert}} = 6$. This is illustrated by Fig. 4, where it is observed that $\text{Conset}_k(\mathbf{A}, \mathbf{b})$ stops growing for $k \geq 6$, that is $\text{Conset}_k(\mathbf{A}, \mathbf{b}) = \text{Conset}_6(\mathbf{A}, \mathbf{b})$ for any $k \geq 6$. One can also notice from Fig. 4c that $C_{\text{lim}} \subset \text{Conset}_{k_{\text{vert}}}(\mathbf{A}, \mathbf{b})$. Note that in this particular example, since $h = 1$, we have $C_{\text{lim}} = \text{cone}(\mathbf{A}_f, \mathbf{b}) = \{c\mathbf{v}_f \mid c \in \mathbb{R}_+\}$, where \mathbf{v}_f is the Frobenius eigenvector of \mathbf{A}^h .

4.2. Special Case

So far it has been assumed that $\text{rank}(\text{conmat}_n(\mathbf{A}, \mathbf{b})) = n$. Based on this assumption, the polyhedrality of the finite controllable subset only depends on the spectrum of \mathbf{A} . In addition, $k_{\text{vert}} \geq n$ for $\text{Conset}_f(\mathbf{A}, \mathbf{b})$. We now point out that in the absence of such an assumption, $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ can depend on the structure of \mathbf{b} and that the vertex number can be less than n . In particular, it will be shown that $k_{\text{vert}} = h$ if $\mathbf{b} \in \mathbb{R}_+^n$ is of a particular structure.

Theorem 4. *Let $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ be irreducible with degree of cyclicity h with $0 \leq h \leq n - 1$. Then, $\text{Conset}_f(\mathbf{A}, \mathbf{b}) = \text{cone}(\text{conmat}_h(\mathbf{A}, \mathbf{b}))$ if $\mathbf{b} \in \text{cone}(\mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1})$, where $\mathbf{v}_{f,i}$, $i = 0, \dots, h - 1$ are the h nonnegative eigenvectors of \mathbf{A}^h .*

PROOF. Assume $\mathbf{b} = \sum_{i=0}^{h-1} c_i \mathbf{v}_{f,i}$ for some $\mathbf{c} \in \mathbb{R}_+^h$. Then, since

$$\mathbf{A}^h \mathbf{b} = \sum_{i=0}^{h-1} c_i \rho(\mathbf{A})^h \mathbf{v}_{f,i} = \rho(\mathbf{A})^h \mathbf{b},$$

it is immediate to see that $\mathbf{A}(\text{conmat}_h(\mathbf{A}, \mathbf{b})) = (\text{conmat}_h(\mathbf{A}, \mathbf{b}))\mathbf{X}$ has a non-negative solution

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \rho(\mathbf{A})^h \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \quad (32)$$

which, in the view of Corollary 2, completes the proof.

For \mathbf{A} primitive (i.e., $h = 1$), this results in the obvious case of $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ being a ray along the Frobenius eigenvector \mathbf{v}_f of \mathbf{A} when $\mathbf{b} = c\mathbf{v}_f$ for any $c \geq 0$.

5. Characterizations of Controllability

Given a cone $C_{\text{obj}} \subseteq \mathbb{R}_+^n$ of control objectives or a subset of \mathbb{R}_+^n , the problem is to investigate whether C_{obj} is contained in $\text{Conset}_f(\mathbf{A}, \mathbf{b})$ or in $\text{Conset}_\infty(\mathbf{A}, \mathbf{b})$. Of particular interest is when $C_{\text{obj}} \subset \mathbb{R}_+^n$ is a polyhedral cone or a polytope. If the control objective cone C_{obj} is not polyhedral then outer approximate it by a polyhedral cone $C_{\text{out}} \subseteq \mathbb{R}_+^n$ such that $C_{\text{obj}} \subset C_{\text{out}}$. Here, it is assumed that the controllability cone or its closure is polyhedral and that its corresponding vertex number or an upper bound of it is known. Hence $\text{Conset}_\infty(\mathbf{A}, \mathbf{b}) = \text{cone}(\mathbf{b} \dots \mathbf{A}^{N-1} \mathbf{b} \mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1})$ for some $N \geq k_{\text{vert}}^\infty$ and/or $\text{Conset}_f(\mathbf{A}, \mathbf{b}) = \text{cone}(\mathbf{b} \dots \mathbf{A}^{N-1} \mathbf{b})$ for some $N \geq k_{\text{vert}}$.

Proposition 4. Let $C_{\text{obj}} = \text{conv}(\mathbf{p}_1, \dots, \mathbf{p}_m)$ or $C_{\text{obj}} = \text{cone}(\mathbf{p}_1, \dots, \mathbf{p}_m)$, where $\mathbf{p}_i \in \mathbb{R}_+^n$, $i = 1, \dots, m$.

- (a) C_{obj} is controllable in finite time if and only if $\mathbf{p} \in \text{Conset}_f(\mathbf{A}, \mathbf{b}), \forall \mathbf{p} \in \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$.

- (b) C_{obj} is controllable in infinite time (to be called *almost controllable*) if and only if $\mathbf{p} \in \text{Conset}_{\infty}(\mathbf{A}, \mathbf{b})$, $\forall \mathbf{p} \in \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$ and $\exists \mathbf{p}' \in \{\mathbf{p}_1, \dots, \mathbf{p}_m\}$ such that $\mathbf{p}' \notin \text{Conset}_f(\mathbf{A}, \mathbf{b})$.

PROOF. The proof is obvious from Definition 10 and considering the fact that a cone can be expressed as a nonnegative combination of its generators.

It is obvious from Proposition 4, that checking for controllability involves checking the following condition for each $i \in \{1, \dots, m\}$:

$$\exists \mathbf{x}_i \in \{\mathbf{z} | \mathbf{M}\mathbf{z} = \mathbf{p}_i, \mathbf{z} \in \mathbb{R}_+^N\}, \quad (33)$$

where $\mathbf{M} \in \mathbb{R}_+^{n \times N}$. Depending on the problem being investigated, either $\mathbf{M} = [\mathbf{b} \dots \mathbf{A}^{N-1}\mathbf{b} \ \mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1}]$ or $\mathbf{M} = [\mathbf{b} \dots \mathbf{A}^{N-1}\mathbf{b}]$.

In general, since $N \geq n$ (see Remark 3 and Remark 4), (33) defines an underdetermined system of equations. It is known that the nonnegative solution of (33) is not unique in general [29, 30], and that uniqueness is guaranteed when the solution is sufficiently sparse [29]. The authors of [31] characterize necessary and sufficient conditions on the polytope $P = \text{conv}(\mathbf{M})$ for uniqueness of the solution, where they prove unique solution exists if and only if P is k -neighborly⁴. In [30, 33], the equivalent of this condition is presented in terms of the null space of \mathbf{M} . In this regard, this problem relates to the *sparse measurement* problem, where it is formulated as reconstructing a nonnegative sparse vector from lower-dimensional linear measurements [34]. The results in this field do not directly apply here as the necessary sparsity condition is usually not met. In addition, we are not interested in finding the sparsest solution of (33), which is normally an NP-hard problem [29].

Proposition 5. Consider index sets $\mathcal{I}_j^i \subset \{1, \dots, N\}$ for $j = 1, \dots, C(N, n)$ with $|\mathcal{I}_j^i| = n$, where $N > n$ is an upper bound to k_{vert} or an upper bound to k_{vert}^{∞} , n is the dimension of space, and $C(N, n)$ is the number of n -combinations

⁴A k -neighborly polytope is a convex polytope in which every set of k or fewer vertices forms a face [32].

of the set $\{1, \dots, N\}$. Let $\mathbf{I}_{\mathcal{I}_j^i}$ denote the submatrix of the identity matrix of dimension N , \mathbf{I}_N that is composed of the columns corresponding to \mathcal{I}_j^i . Then, (33) has a solution \mathbf{x}_i for any $i \in \{1, \dots, m\}$ if and only if

$$\mathbf{X}^i = \left\{ \mathbf{x}_j^i \mid \mathbf{x}_j^i = \mathbf{I}_{\mathcal{I}_j^i} (\mathbf{M} \mathbf{I}_{\mathcal{I}_j^i})^{-1} \mathbf{p}_i, \mathbf{x}_j^i \in \mathbb{R}_+^N, j = 1, \dots, C(N, n) \right\} \quad (34)$$

is a non-empty set.

PROOF. From our assumption we have $\mathbf{p}_i \in \text{cone}(\mathbf{M})$. Since $N > n$, due to the Carathéodory theorem [35], \mathbf{p}_i also lies in at least one simplicial cone generated by n columns of \mathbf{M} . Let $\mathcal{J}^i \in \{1, \dots, N\}$ with $|\mathcal{J}^i| = n$ be an index set composed of the indices of the columns generating this simplicial cone, and let $\mathbf{M}_{\mathcal{J}^i}$ denote the columns of \mathbf{M} corresponding to \mathcal{J}^i . We can then write $\mathbf{p}_i \in \text{cone}(\mathbf{M}_{\mathcal{J}^i})$, which can be expressed as $\mathbf{M} \mathbf{I}_{\mathcal{J}^i} \mathbf{z}^i = \mathbf{p}_i$ having a solution $\mathbf{z}^i \in \mathbb{R}_+^n$. Since \mathbf{M} has full row rank and $\mathbf{I}_{\mathcal{J}^i}$ is full column rank, one obtains $\mathbf{z}^i = (\mathbf{M} \mathbf{I}_{\mathcal{J}^i})^{-1} \mathbf{p}_i$. Finally, we obtain a solution $\mathbf{x}_j^i \in \mathbb{R}_+^N$, where $\mathbf{x}_j^i = \mathbf{I}_{\mathcal{J}^i} \mathbf{z}^i = \mathbf{I}_{\mathcal{J}^i} (\mathbf{M} \mathbf{I}_{\mathcal{J}^i})^{-1} \mathbf{p}_i$.

The converse is proved in a straightforward manner by noticing that every $\mathbf{z} \in \mathbf{X}^i$ satisfies (33).

Remark 5. Let $\mathbf{X}^i = \left\{ \mathbf{x}_1^i, \dots, \mathbf{x}_{q_i}^i \right\}$ for some $q_i \in \mathbb{Z}_+$. It is then clear from the proof of Proposition 5 that the set of solutions of (33) is the convex hull of \mathbf{X}^i , that is, we have for (33) that $\mathbf{x}_i \in \text{conv}(\mathbf{X}^i)$.

Note that even though Proposition 5 provides a method to determine whether $C_{\text{obj}} \subseteq \text{cone}(\mathbf{M})$ by checking inclusion of C_{obj} in any simplicial subcone of $\text{cone}(\mathbf{M})$, the computational complexity of this method can be prohibitive as the check must be conducted for all $C(N, n)$ simplicial subcones in the worst case. A more practical approach is then presented by the following proposition.

Proposition 6. Let

$$\mathbf{M}_f = [\mathbf{b}, \dots, \mathbf{A}^{N-1} \mathbf{b}] \text{ and } \mathbf{M}_\infty = [\mathbf{b}, \dots, \mathbf{A}^{N-1} \mathbf{b}, \mathbf{v}_{f,0}, \dots, \mathbf{v}_{f,h-1}].$$

Define the following optimization problem for each $i \in \{1, \dots, m\}$:

$$\begin{aligned} \min_{\mathbf{x}_i} \mathbf{x}_i^T \mathbb{1} \\ \mathbf{M}\mathbf{x}_i &= \mathbf{p}_i \\ \mathbf{x}_i &\geq 0, \end{aligned} \tag{35}$$

where $\mathbb{1} \in \mathbb{R}^n$ is a vector of ones. We then have the following.

- (a) The optimization problem (35) with $\mathbf{M} = \mathbf{M}_\infty$ has an optimal solution $\mathbf{x}_i^* \in \mathbb{R}_+^N$ if and only if (33) has a solution with $\mathbf{M} = \mathbf{M}_\infty$.
- (b) The optimization problem (35) with $\mathbf{M} = \mathbf{M}_f$ has an optimal solution $\mathbf{x}_i^* \in \mathbb{R}_+^N$ if and only if (33) has a solution with $\mathbf{M} = \mathbf{M}_f$.

PROOF. If (33) has a solution, the set \mathbf{X}^i in (34) is non-empty. As mentioned in Remark 5, the feasible set of 35 is $\text{conv}(\mathbf{X}^i)$. Therefore, the convex optimization problem with linear penalty function converges to the minimum 1-norm solution in the feasible set. The converse is obvious.

Example 5. We conclude this section with an example illustrating the application of Proposition 6. Consider the system matrices of Example 4. Let C_{obj} be the polytope given by

$$C_{\text{obj}} = \left\{ \mathbf{p} \in \mathbb{R}_+^4 \mid \mathbf{p} = \sum_{i=1}^4 \lambda_i \mathbf{p}_i, \lambda_i \geq 0, \sum_{i=1}^4 \lambda_i = 1 \right\},$$

where

$$\begin{aligned} \mathbf{p}_1 &= [1, 3, 1, 1]^T, \quad \mathbf{p}_2 = [1, 3, 4, 3]^T, \\ \mathbf{p}_3 &= [1, 2, 2, 1]^T, \quad \mathbf{p}_4 = [1, 1, 2, 1]^T. \end{aligned}$$

We will now check if the system initially at rest can be steered to any point in C_{obj} in finite time. From example 4, $k_{\text{vert}} = 6$ is known. Thus taking $\mathbf{M} = [\mathbf{b}, \mathbf{A}\mathbf{b}, \dots, \mathbf{A}^5\mathbf{b}]$, we solve for the optimization problem (35) using the Dual-Simplex algorithm implemented in Matlab Optimization Toolbox. The

optimal solutions are obtained as

$$\begin{aligned}\mathbf{x}_1^* &= [0.1209, 0.3735, 0, 0.0078, 0, 0.0001]^\mathrm{T}, \\ \mathbf{x}_2^* &= [2.3460, 0.6165, 0.0876, 0, 0.0003, 0]^\mathrm{T}, \\ \mathbf{x}_3^* &= [0.2989, 0.6982, 0.0473, 0, 0.0003, 0]^\mathrm{T}, \\ \mathbf{x}_4^* &= [0.2517, 0.7798, 0.0071, 0, 0.0003, 0]^\mathrm{T}.\end{aligned}$$

Hence, the vertices of C_{obj} can be reached from the origin in finite number of steps using nonnegative inputs, which are determined by the solution vectors \mathbf{x}_i^* . Moreover, since $k_{\text{vert}} = 6$, every vertex of C_{obj} can be reached in at most 6 steps from the origin. Since C_{obj} is the convex hull of its vertices, we can conclude that any point $\mathbf{p} = \sum_{i=1}^4 \lambda_i \mathbf{p}_i \in C_{\text{obj}}$ can be reached from the origin in at most 6 steps using the input sequence $\mathbf{u}^* = \sum_{i=1}^4 \lambda_i \mathbf{x}_i^*$.

6. Conclusion

We discussed a new view of the controllability problem for linear time-invariant positive systems that is more interesting for practical applications than the classical view. The controllability was defined as the ability to drive the system initially at origin to a certain target subset of \mathbb{R}_+^n using nonnegative inputs. To this end, we discussed the geometry of controllable subsets and developed sufficient and necessary conditions for polyhedrality of such subsets. We showed that when the controllability matrix of the system is of full rank, those conditions solely depend on the spectrum of \mathbf{A} . In addition, it was shown that the controllable subset may keep growing for more than n steps, where n is the dimension of the system. We then proposed a numerical method to check for controllability of a linear positive system with respect to a certain objective set.

In this paper, we have focused on the single input case, where $\mathbf{b} \in \mathbb{R}_+^n$. The controllability problem for the multi-input case is an interesting problem as the results developed here are not directly applicable. The main issue, as noted in [28], is that the direct sum of two non-polyhedral cones may result in

a polyhedral cone. Therefore, one cannot apply the results of this paper to a set of system $(\mathbf{A}, \mathbf{b}_i)$ separately, with \mathbf{b}_i being a column of \mathbf{B} .

It is also of interest to investigate the geometry of controllable subsets when the controllability matrix is not of full rank. As far as the authors of this paper know, this is still an open issue.

Appendix A. Positive Matrices

For completeness, we report Theorem 5 of [27] here. In this theorem, Q denotes the set of all real polynomials of the form $c_n x^n - \sum_{i=0}^{n-1} c_i x^i$, where $n \geq 1$, $c_n > 0$, and $c_i \geq 0$ for all i .

Theorem 5 ([27, Th. 5]). *Let $\{a_1, \dots, a_k\}$ be given complex numbers, and let $P(x)$ be the polynomial $x^k - a_1 x^{k-1} - \dots - a_k$. Then conditions (A), (B) and (C) below are equivalent:*

- (A) *Any infinite sequence $(u_n)_{n \geq 0}$ of complex numbers which satisfies the recursion $u_{n+k} = a_1 u_{n+k-1} + a_2 u_{n+k-2} + \dots + a_k u_n$ for $n \geq 0$, also satisfies a recursion with nonnegative coefficients.*
- (B) *The polynomial $P(x)$ divides a polynomial in Q .*
- (C) *In case the polynomial $P(x)$ has a positive root r , then all conditions (1)-(4) below are satisfied:*
 - (C1) $r \geq |\alpha|$ for any root α of $P(x)$.
 - (C2) if $\alpha = r$ for some root α of $P(x)$, then α/r is a root of unity.
 - (C3) all roots $P(x)$ with absolute value r are simple.
 - (C4) if $P(r) = P(r\epsilon) = 0$, where $\epsilon^k = 1$ with $k \geq 1$ minimal, then $P(x)$ has no roots of the form $s\omega$ where $0 < s < r$ and $\omega^k = 1$.

Lemma 1. *Let $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ be irreducible with cyclicity index h and let $\mathbf{b} \in \mathbb{R}_+^m$. Define $C_{\lim} = \text{cone}(\mathbf{A}_{f,0}\mathbf{b}, \dots, \mathbf{A}_{f,h-1}\mathbf{b})$, where $\mathbf{A}_{f,i} = \lim_{k \rightarrow \infty} \frac{\mathbf{A}^{kh}}{\rho(\mathbf{A})^{kh}} \mathbf{A}^i$, for $i = 0, \dots, h-1$. Let the nonnegative vectors $\mathbf{v}_{f,i}$, $i = 0, \dots, h-1$ of Proposition 2 be the h distinct nonnegative eigenvectors of \mathbf{A}^h associated with the Perron root of $\rho(\mathbf{A})^h$. It then holds that $C_{\lim} \subseteq \text{cone}(\mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1})$,*

PROOF. Since \mathbf{A} is irreducible, there exists a monomial matrix $\mathbf{S} \in \mathbb{R}_+^{n \times n}$ [9,

Th. 2.2.33], such that

$$\hat{\mathbf{A}} = \mathbf{S}^T \mathbf{A} \mathbf{S} = \begin{bmatrix} 0_{n_1} & A_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0_{n_{h-1}} & A_{h-1} \\ A_h & 0 & \dots & \dots & 0_{n_h} \end{bmatrix}, \quad \hat{\mathbf{b}} = \mathbf{S}^T \mathbf{b} \quad (\text{A.1})$$

where $0_{n_i} \in \mathbb{R}^{n_i \times n_i}$, $i \in \mathbb{N}$ are square blocks with $\sum_{i=1}^h n_i = n$, and where A_i has

no zero rows or columns with $L_1 = \prod_{i=1}^h A_i$ being an irreducible matrix. Then

we have $\hat{\mathbf{A}}^h = \text{diag}(L_1, \dots, L_h)$, where $L_k = \prod_{i=k}^h A_i \prod_{j=1}^{\text{mod}(h+k-1, h)} A_j$ is a primitive matrix of dimension $n_k \times n_k$ with Perron root $\rho(\mathbf{A})^h$. Define the matrix $\hat{\mathbf{A}}_{f,i} = \lim_{p \rightarrow \infty} \frac{\hat{\mathbf{A}}^{ph}}{\rho^{ph}} \hat{\mathbf{A}}^i$ for $i = 0, \dots, h-1$. Since L_i , $i = 1, \dots, h$ is primitive, it follows from [9, Th. 2.4.1] that

$$\hat{\mathbf{A}}_{f,0} = \begin{bmatrix} \mathbf{x}_1^1 & \dots & \mathbf{x}_1^{n_1} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \mathbf{x}_2^1 & \dots & \mathbf{x}_2^{n_2} & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{x}_h^1 & \dots & \mathbf{x}_h^{n_h} \end{bmatrix}, \quad (\text{A.2})$$

where $\mathbf{x}_i^k = c_i^k \mathbf{x}_i$ with c_i^k , $k = 1, \dots, n_i$, being some nonnegative scalars and with $\mathbf{x}_i \in \mathbb{R}_{s+}^{n_i \times n_i}$ being the Frobenius eigenvector of L_i . Note that due to the block structure of $\hat{\mathbf{A}}$, $\hat{\mathbf{A}}_{f,i}$ retains the same structure as $\hat{\mathbf{A}}_{f,0}$ up to a scaled permutation of its columns for $i = 1, \dots, h-1$. Hence, we have $\hat{\mathbf{A}}_{f,i} \hat{\mathbf{b}} \in \text{cone}(\mathbf{C})$, where

$$\mathbf{C} = \begin{bmatrix} \mathbf{x}_1 & 0 & \dots & 0 \\ 0 & \mathbf{x}_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \mathbf{x}_h \end{bmatrix}.$$

In the original coordinates, we have $\mathbf{A}_{f,i}\mathbf{b} \in \text{cone}(\mathbf{SC})$. Clearly, since the columns of \mathbf{C} are the nonnegative eigenvectors of $\hat{\mathbf{A}}^h$ and since \mathbf{S} is monomial, we have $\mathbf{SC} = [\mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1}]$, where $\mathbf{v}_{f,i} \in \mathbb{R}_+^{n \times n}$ is the $(i+1)$ -th nonnegative eigenvector of \mathbf{A}^h for $i = 0, \dots, h-1$. This proves that $\text{cone}(\mathbf{A}_{f,0}\mathbf{b} \dots \mathbf{A}_{f,h-1}\mathbf{b}) \subseteq \text{cone}(\mathbf{v}_{f,0} \dots \mathbf{v}_{f,h-1})$.

Lemma 2. *Let $\mathbf{A} \in \mathbb{R}_+^{n \times n}$ be irreducible with degree of cyclicity h with $1 \leq h \leq n$. Let \mathbf{A} be decomposed as $\mathbf{A} = \mathbf{S}\text{diag}(\mathbf{A}_1, \mathbf{A}_2)\mathbf{S}^{-1}$, where $\text{spec}(\mathbf{A}_1) = \sigma^\rho(\mathbf{A})$ and $\text{spec}(\mathbf{A}_2) = \sigma^-(\mathbf{A})$. Let $\sigma^0 \subseteq \sigma^\rho(\mathbf{A}_2)$ be the set of all eigenvalues of \mathbf{A}_2 whose modulus is $\rho(\mathbf{A}_2)$ and whose polar angle is a rational multiple of 2π . Then, there exists a minimal $M \in \mathbb{Z}_+$ such that*

$$\sigma_0 \subseteq \left\{ \lambda \in \text{spec}(\mathbf{A}_2) \mid \lambda = \rho(\mathbf{A}_2) \exp\left(\frac{2\pi k}{Mh}i\right), k = 0, \dots, Mh-1 \right\}, \quad (\text{A.3})$$

or, equivalently, there exists a minimal $M \in \mathbb{Z}_+$ such that the eigenvalues of $\mathbf{A}_2/\rho(\mathbf{A}_2)$ with unit modulus whose argument are a rational multiple of 2π are among the Mh -th roots of unity.

PROOF. Let δ^0 be a set of $n_{\delta^0} \in \mathbb{Z}_+$ members of σ^0 with the property that the difference between the polar angle of no two members of δ^0 is an integer multiple of $2\pi/h$, or formally we define $\delta^0 = \{\lambda_1, \dots, \lambda_{n_{\delta^0}} \in \sigma^0 \mid \arg(\lambda_i) - \arg(\lambda_j) \neq 2z\pi/h, i \neq j, z \in \mathbb{Z}\}$. For $\lambda_j \in \delta^0$, $j = 1, \dots, n_{\delta^0}$, let $\arg(\lambda_j) = \frac{2\pi p_j}{q_j}$. Define the sets $\sigma_j^0 \subset \sigma^0$ for $j = 1, \dots, n_{\delta^0}$ as

$$\sigma_j^0 = \left\{ \lambda \in \text{spec}(\mathbf{A}_2) \mid \lambda = \rho(\mathbf{A}_2) \exp\left((k/h + p_j/q_j)2\pi i\right), k = 0, \dots, h-1 \right\},$$

or equivalently using the notation $s_{j,k} \equiv kq_j + hp_j \pmod{hq_j}$,

$$\sigma_j^0 = \left\{ \lambda \in \text{spec}(\mathbf{A}_2) \mid \lambda = \rho(\mathbf{A}_2) \exp\left(\frac{s_{j,k}}{hq_j}2\pi i\right), k = 0, \dots, h-1 \right\}.$$

It is clear that $\sigma_1^0, \dots, \sigma_{n_{\delta^0}}^0$ are mutually disjoint. In addition, since the eigenvalues of \mathbf{A} are invariant under polar rotation of $2k\pi/h$ for any $k \in \mathbb{Z}$, we have $\sigma^0 = \bigcup_{j=1}^{n_{\delta^0}} \sigma_j^0$. Noting that $0 \leq s_{j,k} \leq hq_j - 1$ for $k = 0, \dots, h-1$ and for $j = 1, \dots, n_{\delta^0}$, one observes that σ_0 has the form proposed in (A.3) by choosing $M = \text{lcm}(q_1, \dots, q_{n_{\delta^0}})$.

References

References

- [1] W. Leontief, Input-Output Economics, 2nd Edition, Oxford University Press, 1986.
- [2] J. A. Jacquez, Compartmental Analysis in Biology and Medicine, 3rd Edition, BioMedware, 1996.
- [3] E. D. Sontag, Structure and stability of certain chemical networks and applications to the kinetic proofreading model of t-cell receptor signal transduction, IEEE Transactions on Automatic Control 46 (7) (2001) 1028–1047. doi:10.1109/9.935056.
- [4] J. van den Hof, Positive linear observers for linear compartmental systems, SIAM Journal on Control and Optimization 36 (2) (1998) 590–608. doi:10.1137/S036301299630611X.
- [5] W. M. Haddad, V. Chellaboina, Q. Hui, Nonnegative and compartmental dynamical systems, Princeton University Press, Princeton, 2010.
- [6] Y. Zeinaly, B. De Schutter, H. Hellendoorn, An integrated model predictive scheme for baggage-handling systems: Routing, line balancing, and empty-cart management, IEEE Transactions on Control Systems Technology 23 (4) (2015) 1536–1545. doi:10.1109/TCST.2014.2363135.
- [7] R. Shorten, F. Wirth, D. Leith, A positive systems model of tcp-like congestion control: asymptotic results, IEEE/ACM Transactions on Networking 14 (3) (2006) 616–629. doi:10.1109/TNET.2006.876178.
- [8] A. Berman, M. Neumann, R. J. Stern, Nonnegative Matrices in Dynamical Systems, John Wiley & Sons, Inc., 1989.
- [9] A. Berman, R. J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, academic press, 1979.

- [10] C. Davis, Theory of positive linear dependence, *American Journal of Mathematics* 76 (4) (1954) 733–746.
- [11] D. Gale, Convex polyhedral cones and linear inequalities, in: Tj. C. Koopmans (Ed.), *Activity analysis of production and allocation*, Wiley & Sons, New York, 1951, pp. 287–297.
- [12] J. S. Vandergraft, Spectral properties of matrices which have invariant cones, *SIAM Journal on Applied Mathematics* 16 (6) (1968) 1208–1222.
- [13] L. Farina, S. Rinaldi, *Positive Linear Systems: Theory and Applications*, John Wiley & Sons, Inc., 2000.
- [14] P. G. Coxson, H. Shapiro, Positive input reachability and controllability of positive systems, *Linear Algebra and its Applications* 94 (1987) 35–53. doi:10.1016/0024-3795(87)90076-0.
- [15] R. Bru, S. Romero, E. Sánchez, Canonical forms for positive discrete-time linear control systems, *Linear Algebra and its Applications* 310 (1-3) (2000) 49–71. doi:10.1016/S0024-3795(00)00044-6.
- [16] D. G. Luenberger, *Introduction to Dynamic Systems: Theory, Models & Applications*, John Wiley & Sons, Inc., 1979.
- [17] M. E. Valcher, Controllability and reachability criteria for discrete-time positive systems, *International Journal of Control* 65 (3) (1996) 511–536.
- [18] M. P. Fanti, B. Maione, B. Turchiano, Controllability of multi-input positive discrete-time systems, *International Journal of Control* 51 (6) (1990) 1295–1308. doi:10.1080/00207179008934134.
- [19] C. Guiver, D. Hodgson, S. Townley, Positive state controllability of positive linear systems, *Systems & Control Letters* 65 (2014) 23–29. doi:10.1016/j.sysconle.2013.12.002.

- [20] Z. Bartosiewicz, Linear positive control systems on time scales; controllability, *Mathematics of Control, Signals, and Systems* 25 (3) (2013) 327–343. doi:10.1007/s00498-013-0106-6.
- [21] L. Caccetta, V. G. Rumchev, A survey of reachability and controllability for positive linear systems, *Annals of Operations Research* 98 (1-4) (2000) 101–122. doi:10.1023/A:1019244121533.
- [22] T. Kaczorek, Some recent developments in positive and compartmental systems, in: *SPIE 5484, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*, Vol. 5484, 2004, pp. 1–18. doi:10.1117/12.568841.
- [23] T. Kaczorek, *Positive 1D and 2D Systems*, Springer, 2002.
- [24] H. Minc, *Nonnegative Matrices*, Wiley, 1989.
- [25] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [26] P. G. Coxson, L. C. Larson, H. Schneider, Monomial patterns in the sequence $a^k b$, *Linear Algebra and its Applications* 94 (1987) 89–101. doi:10.1016/0024-3795(87)90080-2.
- [27] M. Roitman, Z. Rubinstein, On linear recursions with nonnegative coefficients, *Linear Algebra and its Applications* 167 (1992) 151–155. doi:10.1016/0024-3795(92)90344-A.
- [28] L. Benvenuti, L. Farina, The geometry of the reachability set for linear discrete-time systems with positive controls, *SIAM. Journal on Matrix Analysis & Applications* 28 (2) (2006) 306–325. doi:10.1137/040612531.
- [29] D. L. Donoho, J. Tanner, Sparse nonnegative solution of underdetermined linear equations by linear programming, *Proceedings of the National Academy of Sciences of the United States of America* 102 (27) (2005) 9446–9451.

- [30] M. Wang, W. Xu, A. Tang, A unique “nonnegative” solution to an under-determined system: From vectors to matrices, *IEEE Transactions on Signal Processing* 59 (3) (2011) 1007–1016. doi:10.1109/TSP.2010.2089624.
- [31] D. L. Donoho, High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension, *Discrete & Computational Geometry* 35 (4) (2006) 617–652. doi:10.1007/s00454-005-1220-0.
- [32] B. Grunbaum, G. M. Ziegler, *Convex Polytopes*, 2nd Edition, Graduate Texts in Mathematics, Springer-Verlag, 2003.
- [33] D. L. Donoho, J. Tanner, Counting the faces of randomly-projected hypercubes and orthants, with applications, *Discrete & Computational Geometry* 43 (3) (2010) 522–541. doi:10.1007/s00454-009-9221-z.
- [34] M. A. Khajehnejad, A. G. Dimakis, W. Xu, B. Hassibi, Sparse recovery of nonnegative signals with minimal expansion, *IEEE Transactions on Signal Processing* 59 (1) (2011) 196–208. doi:10.1109/TSP.2010.2082536.
- [35] I. Bárány, R. Karasev, Notes about the carathéodory number, *Discrete & Computational Geometry* 48 (3) (2012) 783–792. doi:10.1007/s00454-012-9439-z.